Check for updates

RESEARCH ARTICLE

REVISED SARS-CoV-2 epidemic in India: epidemiological features and *in silico* analysis of the effect of interventions [version 2; peer review: 2 approved]

Archisman Mazumder [iD] [1], Mehak Arora [iD] [1], Vishwesh Bharadiya[1], Parul Berry [iD] [1], Mudit Agarwal[1], Priyamadhaba Behera [iD] [2], Hemant Deepak Shewade [iD] [3,4], Ayush Lohiya[5], Mohak Gupta [iD] [1], Aditi Rao[1], Giridara Gopal Parameswaran[1]

[1]All India Institute of Medical Sciences, New Delhi, Delhi, India
[2]All India Institute of Medical Sciences, Raebareli, Uttar Pradesh, India
[3]Operational Research, The Union South-East Asia Office, New Delhi, Delhi, India
[4]Centre For Operational Research, International Union Against Tuberculosis and Lung Disease (The Union), Paris, France
[5]Super Specialty Cancer Institute & Hospital, Lucknow, India

Abstract

Background: After SARS-CoV-2 set foot in India, the Government took a number of steps to limit the spread of the virus in the country. This included restricted testing, isolation, contact tracing and quarantine, and enforcement of a nation-wide lockdown starting 25 March 2020. The objectives of this study were to i) describe the age, gender distribution, and mortality among COVID-19 patients identified till 14 April 2020 and predict the range of contact rate; and ii) predict the number of COVID-19 infections after 40 days of lockdown.

Methods: We used a cross-sectional descriptive design for the first objective and a susceptible-infected-removed model for *in silico* predictions. We collected data from government-controlled and crowdsourced websites.

Results: Studying age and gender parameters of 1161 Indian COVID-19 patients, the median age was 38 years (IQR, 27-52) with 20-39 year-old males being the most affected group. The number of affected patients were 854 (73.6%) men and 307 (26.4%) women. If the current contact rate continues (0.25-27), India may have 110460 to 220575 infected persons at the end of 40 days lockdown.

Conclusion: The disease is majorly affecting a younger age group in India. Interventions have been helpful in preventing the worst-case scenario in India but will be unable to prevent the spike in the number of cases.

Keywords
COVID-19, Coronavirus, SIR model, India

Open Peer Review

Reviewer Status ✔✔

	Invited Reviewers	
	1	**2**
version 2 (revision) 29 Jun 2020		
version 1 30 Apr 2020	✔ report	✔ report

1 **Sarika Palepu** [iD] , Vir Chandra Singh Garhwali Government Institute of Medical Sciences and Research, Srinagar, India

2 **Soumya Swaroop Sahoo**, All India Institute of Medical Sciences, Bathinda, India

Any reports and responses or comments on the article can be found at the end of the article.
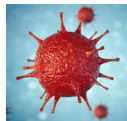
This article is included in the Disease Outbreaks gateway.
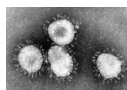
This article is included in the Coronavirus collection.

Corresponding author: Priyamadhaba Behera (priya.madhaba@gmail.com)

Author roles: Mazumder A: Formal Analysis, Methodology, Writing – Original Draft Preparation; **Arora M**: Formal Analysis, Methodology, Writing – Original Draft Preparation; **Bharadiya V**: Formal Analysis, Methodology, Writing – Original Draft Preparation; **Berry P**: Formal Analysis, Methodology, Writing – Original Draft Preparation; **Agarwal M**: Validation, Writing – Review & Editing; **Behera P**: Conceptualization, Supervision, Writing – Review & Editing; **Shewade HD**: Conceptualization, Writing – Review & Editing; **Lohiya A**: Conceptualization, Writing – Review & Editing; **Gupta M**: Validation, Writing – Review & Editing; **Rao A**: Validation, Writing – Review & Editing; **Parameswaran GG**: Conceptualization, Writing – Review & Editing

How to cite this article: Mazumder A, Arora M, Bharadiya V *et al.* **SARS-CoV-2 epidemic in India: epidemiological features and *in silico* analysis of the effect of interventions [version 2; peer review: 2 approved]** F1000Research 2020, **9**:315 https://doi.org/10.12688/f1000research.23496.2

First published: 30 Apr 2020, **9**:315 https://doi.org/10.12688/f1000research.23496.1

Introduction

Since December 2019, SARS-CoV-2, a novel virus of the *Coronaviridae* family of RNA viruses, has caused a widespread outbreak of the disease, now known as COVID-19, and was declared to be a pandemic by the World Health Organization (WHO) on March 11 2020[1–3]. Human to human transmission occurs primarily through close-contact with the infected person, through fomites in the immediate surroundings of the infected person and via droplets of respiratory secretion, although there is limited evidence pointing to a possibility of airborne and faeco-oral transmission as well[4–7]. According to a few case studies, transmission may also occur via viral shedding in "pre-symptomatic" individuals during the incubation period[8,9].

The incubation period for COVID-19 is thought to be within 14 days of exposure, with a median incubation period of 4–5 days[4,10,11]. Globally, the median age of patients affected by COVID-19 is 47 years with the most common clinical findings being fever and cough[4,12]. About 18% of patients develop shortness of breath[4]. Severe disease (including dyspnea defined as a respiratory rate of 30/min, blood oxygen saturation of 93%, a partial pressure of arterial oxygen to fraction of inspired oxygen ratio<300, and/or lung infiltrates>50% within 24 to 48 hours) has been reported to occur in 14% of elderly patients with pre-existing chronic diseases[13]. Critical disease requiring intensive care unit admission has been reported in 5% of patients, and overall case-fatality rate is 2.3%[13]. Currently, there are no approved treatments for COVID-19 and clinical trials, such as the WHO SOLIDARITY trial, are underway to evaluate the effectiveness of drugs like lopinavir-ritonavir, remdesivir, hydroxychloroquine and azithromycin[3,14].

India reported its first case of COVID-19 on 30 January 2020; a medical student who had travelled from Wuhan, China, the then epicentre of COVID-19[15]. On 15th March 2020, India sealed its borders and stopped all international flights, meaning all initial imported cases in India arrived before 15 March 2020. According to the data available in the public domain, as of 12 April 2020, India had 8606 cases (both imported cases and due to person-to-person transmission). According to the Indian Council of Medical Research (the apex national medical research body), as of 31st March 2020, community transmission had not started then and India was in category 2 of WHO classification for transmission patterns, i.e. sporadic cases without evidence of community transmission[16]. Recent evidence suggests 104 (1.8%) of the 5,911 patients with severe acute

respiratory infection who tested positive for COVID-19, and 40 (39.2%) COVID-19 cases did not report any history of contact with a known case or international travel. This indicates imminent community transmission in the near future[17].

Since the beginning of the outbreak in India, there have been a number of interventions done by both state level and central level governments (Figure 1). These included restricting the inflow of international passengers, self-quarantine measures, directives on testing and management strategies, and in-country travel restrictions. During this period, the testing strategy also changed from being focused on foreign travel and contact history initially to including all individuals with severe acute respiratory illness and symptomatic health care providers Gradually there were more social distancing measures in March, which were then followed by state-wise lockdowns, ultimately culminating in a nation-wide 21 day lockdown from 25 March 2020. As of 14th April, the lockdown had been extended until 3rd May 2020 (40 days).

In addition to the epidemiological parameters of COVID-19 in India, the other important question in the current scenario concerns the mathematical parameters of the initial spread of COVID-19 in India, and what are the epidemiological aspects that can predict this spread. We acknowledge there are certain difficulties in making precise calculations due to the rapidly changing dynamicity of the epidemic in the early stages, limited availability of data in the public domain, and limited testing capacity. Nonetheless, mathematical models with reasonable assumptions based on available information can help in analysis of the currently available data to provide important insights for guiding public health interventions[18]. The most basic of these models is the susceptible-infected-removed (SIR) model[19,20], which we have used in the current Indian scenario to determine the range in which contact rate β lies, calculate the range of the current reproduction number, Rt, and predict the number of COVID-19 infections at the end of the 40 day lock down period.

Methods

Epidemiological descriptive analysis of patients

This was a cross-sectional descriptive analysis of the laboratory confirmed COVID-19 patient-wise data collected from a crowdsourced database (https://www.covid19india.org; includes data from state government and central government agencies). The data was taken for cases confirmed up to 14 April 2020, 7:20 PM Indian standard time or earlier.

Data analysis was done in regard to the age distribution, status of patients and gender distribution using Microsoft Office Excel 2007 (Microsoft, Redmond, WA, USA). Fatality rate in any category was found by dividing the number of deaths in the category by the number of affected individuals of that category.

Mathematical modelling

The SIR model[19,20] divides the (fixed) population of N individuals into three "compartments", which vary as a function of time (for purpose of this study, we have not included vital dynamics
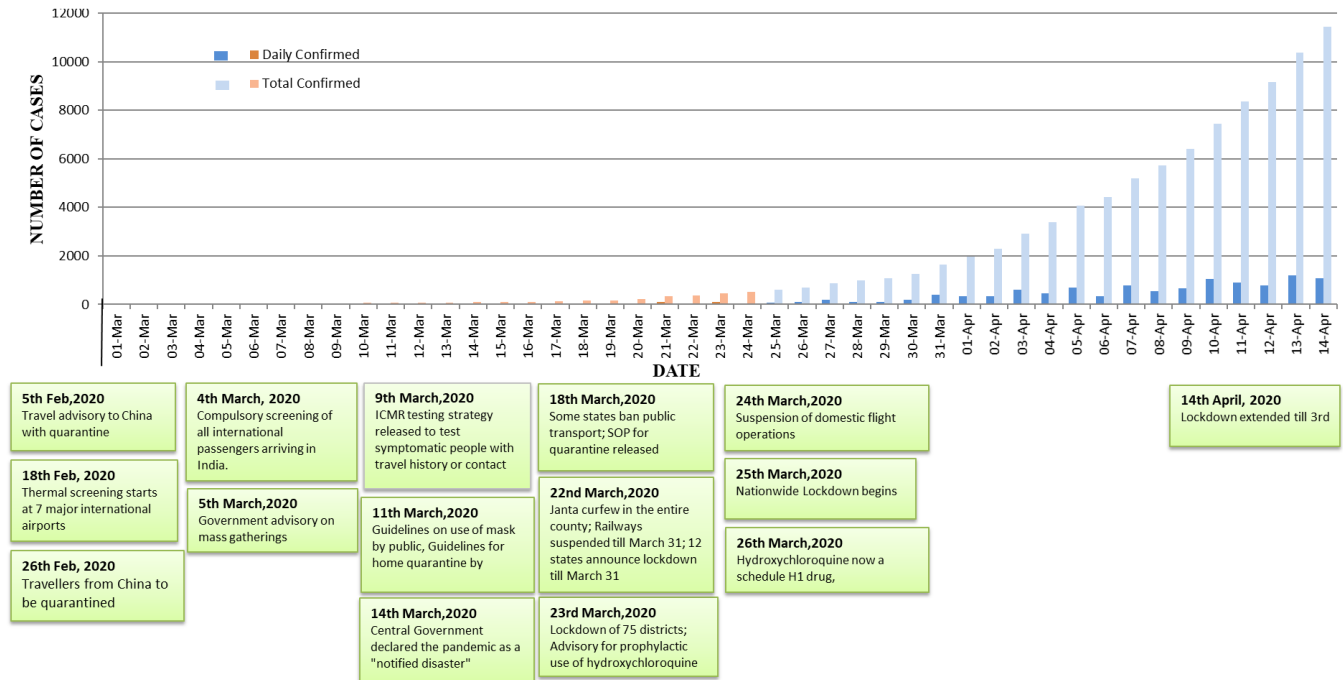
Figure 1. Cumulative cases of COVID-19 cases in India from 1st March 2020 to 14th April 2020. Data Source: ICMR-NIE and MoHFW (www.mohfw.gov.in and www.covidindiaupdates.in). Orange, before lockdown; blue, after lockdown began.

like birth and death rate because in COVID-19, the duration of infection is much less than the lifetime of an individual and hence they would not significantly affect the results):

- S(t) - S(t) are those susceptible but not yet infected with the disease (in a novel disease like COVID-19, the entire population is assumed to be susceptible as there is no pre-existing immunity);

- I(t) - I(t) is the number of infectious individuals;

- R(t) - R(t) are those individuals who have been removed from the infected population (includes those who have recovered from the disease and also deaths).

The SIR model describes the change in the population within each of these compartments in terms of two parameters, β and γ (Figure 2). β describes the effective *contact rate* of the disease: a susceptible individual comes into contact with an infectious individual and acquires the infection. This parameter considers both the number of people contacted per unit time, and the effectiveness of transmission in each contact. It reflects the force of infection of the disease and helps us understand at what rate the epidemic is progressing. γ is the mean *removal rate*: in our model, it is calculated using the removed cases against new cases on a daily basis.

β and γ are useful in the SIR model using the following differential equations:

$$dS/dt = -\beta SI/N \qquad [1]$$

$$dI/dt = \beta SI/N - \gamma I \qquad [2]$$

$$dR/dt = \gamma I \qquad [3]$$

Comparing the equation from the SIR model and the general equation of exponential growth,

$$\beta = e^{r} + \gamma - I \qquad [4]$$

where r is the growth rate of the exponential curve. The Susceptible-Infectious-Removed-Susceptible (SIRS) model because reinfection has been rarely reported for COVID-19.

We have used the new cases of COVID-19 daily data available from 24th March to 13th April 2020, to estimate the two parameters (assuming a lag period of 11 days), β and R_t (time varying reproductive number) with the help of SIR model (www.statista.com used for extraction of variables). We assumed that the recovery rate γ would remain constant for the population. The removal rate followed a normal distribution, and the mean was calculated with the data available from 1 March to April 4 2020[21]. Since the effect of interventions would reflect in the contact rate β, we then took the value of γ to be constant equal to the mean (0.103) and ran the SIR model multiple times by varying the value of β, and comparing the trends with the real data. We assumed uniform transmission for all the simulations.

Figure 2. Flow diagram of transitions in susceptible-infected-removed (SIR) Model.

We plotted the trend line for the real data using Microsoft Office Excel 2007, and used the equation of the curve to find out the trend of β in India in the present-day scenario by comparing it to equation 4. We then used the present trends of β to estimate the expected number of infections at the end of 40 days lockdown.

Ethics

Anonymized data available data in the public domain was used for analysis. Ethical approval was not required.

Results

Epidemiological descriptive analysis of patients

There were 10939 total patients recorded; age and gender data were available for 1161 (10.61%). Of 1161 patients, 854 (73.6%) were men and 307 (26.4%) were women. Median age

was 38 years (IQR, 27–52). Nearly half of the patients (43.2%) were in the age group of 20–39 years. The median age of women was 40 years (IQR, 24–56.5) and men was 38 years (IQR, 27–51) (Table 1).

Out of 1161 patients, 1038 hospitalizations and 29 deaths were documented. The majority of them (65.5%) were in the 60–79 year age group. The median age of deceased patients was 67 years (IQR, 57–71). The mortality rate in male patients was 2.3% and for female patients was 2.9%. Overall case fatality was 2.5%. There was a 10.6% mortality in the 60–79 years age group and 16.7% in individuals aged equal to or above 80 years (Table 2).

In total, 94 patients had recovered. Among the recovered patients, 47 (50.0%) were in the 20–39 year age group,

Table 1. Distribution of COVID-19 patients in India confirmed until 14 April 2020 across different age groups and gender (n=1161).

Age group	Men		Women	
	N	%(normalized value)*	N	%(normalized value)*
<20years	66	5.68(0.31)	43	3.70(0.22)
20–39 years	391	33.68(1.94)	110	9.47(0.61)
40–59 years	274	23.60(2.13)	86	7.41(0.70)
60–79 years	116	9.99(2.20)	63	5.43(1.17)
>=80 years	7	0.60(1.40)	5	0.43(0.81)
Total	**854**		**307**	

*Normalized values have been calculated by dividing the percentage of patients in each category by the percentage of Indian population in that category.

Table 2. Distribution of COVID-19 patients of India with age, gender and status of disease updated till 14 April 2020 (n=1161).

Variables	Deceased	Recovered	Hospitalized*
AGE			
<20years	1	8	100
20–39 years	2	47	452
40–59 years	5	22	333
60–79 years	19	14	146
>=80 years	2	3	7
GENDER			
Male	20	66	768
Female	9	28	270
TOTAL	**29**	**94**	**1038**

*Hospitalized include patients in facility isolation, in ICU or under medical care in hospital

22(23.4%) in the 40–59 year age group. Median age for recovered patients was 35 years (IQR, 24–53).

Mathematical modelling

After running multiple simulations using the SIR model, all assuming different values of β we found that the value of β in the current Indian scenario calculated from the trendline of real data lies around 0.272, which is also visible in the graph of real-time active cases lying between β=0.25 and β=0.28 curves (Figure 3).

Also visible in the graph, is the real data line shifting from 0.24 to 0.28 from day 8. Also, as $R_t = β/γ$, the value of the reproductive number is found to be varying between 2.5–2.8 in the Indian scenario until 13th April 2020. Considering the present trend of β, the number of infected persons after the 40 day lockdown period may vary from 110,460 to 220,575.

Discussion

To the best of our knowledge, this is the first study to predict the number of COVID-19 infected persons in India after the 40 days of lockdown. We have used crowd sourced data due to lack of availability of official data. We could not include all the cases in the epidemiological analysis due to unavailability of the demographic details of all the patients. Limitations notwithstanding, there were three key findings.

First, the median age of affected individuals is lower than reported in other countries. In study cohorts of Wuhan, the median age of affected patients ranges from 49–56 years[4,22,23]. In previous studies on COVID-19, it has been established that risk increases with age and comorbidities[13,22,23]. The broad based nature of India's population pyramid means there are more people in the younger age group and very few people in the ≥80 years age group. Hence, because of more percentage of the population in the younger age group, they are more likely to be infected. Another reason is that as there is limited evidence of community spread in the epidemic in India so far; it has been reported to be driven by imported cases who mostly belong to the younger age group. In 2018, Indian residents between 35 and 49 years of age took the most holidays outside the nation[24]. Hence, in the early part of the epidemic, younger people would be more affected as they would constitute more of the imported cases. On normalizing the percentage of patients in each age group with the corresponding percentage representation of the population as shown in Table 1, we observed that the highest normalized ratio of number of patients is in the 60–79 years age group category. Interestingly, according to this analysis, men in the 60–79 years age group are affected more than the ≥80 years age group. This is something that has not been reported until now and it has to be seen whether this changes as the number of cases in India grows. This is new to the literature of COVID-19 and needs to be studied further.
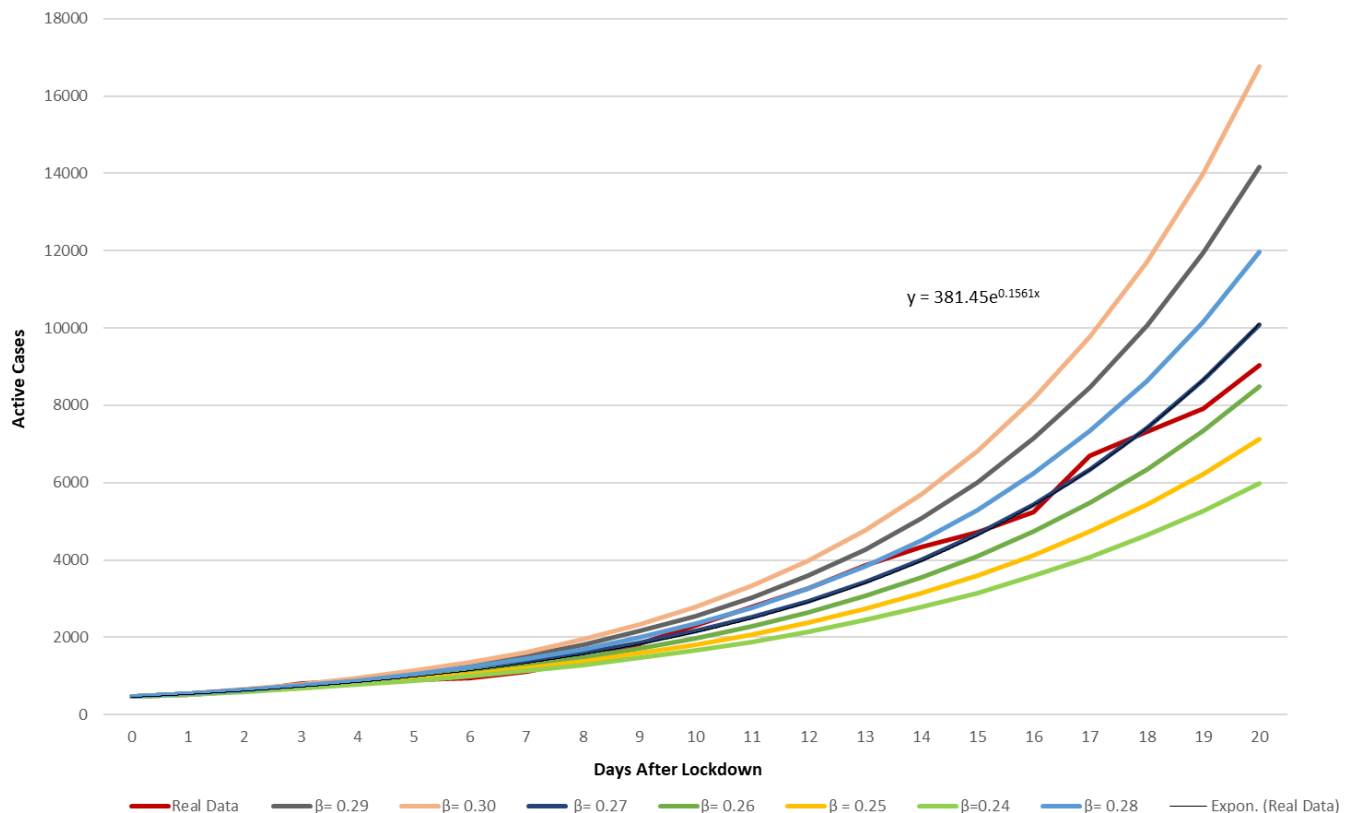

Figure 3. Estimation of β from Indian data. Datasource: https://www.statista.com/statistics/1104054/india-coronavirus-covid-19-daily-confirmed-recovered-death-cases/.

Second, we found a case fatality rate of 2.5%, which is lower than countries like Italy[25]. The reason for this could be that Italy is in the later phases of the epidemic with widespread community transmission and high mortality due to overburdening of the health system. This has also led to focussing testing on severe patients and possible advise to suspected patients with mild symptoms to stay at home. In China, the case fatality rate was found as 2.3%, 14.8% in the above 80 years population, and 8.0% in the 70–79 years population[13]. Our estimate gives a slightly higher value, which may be due a smaller sample size or can also be because mild cases of COVID-19 have so far been missed due to restricted testing strategy during the initial stage.

Third, from the SIR model, the values of β and the trend for R_t shows that the interventions which were put in place by the Indian government starting from mid-March were partially effective, preventing the scenario where R_t can reach even more than four[26]. The increase in β from day eight is probably due to the identification of a cluster in Delhi[27]. Additionally, Indian Council of Medical Research recently updated the testing strategy, which might have increased the detection of the positive cases significantly. According to a study in the early days of the epidemic, Wuhan city and Hubei province reported R_t between 1.85 and 4.46, which aligns with our study findings[26]. All over China, the R_t varied from 1.23 to 5.77. South Korea, which has high population density like India, had a decreasing trend of R_t from 9.72 on 20 February to 1.50 on 7 March. This indicates that the interventions have been helpful in preventing the worst case scenario in India but is unable to prevent the spike in number of cases[27,28].

At the end of the 40 days lockdown, with the range of β at the time of study, India is likely to have a significant number of infected persons (110,460 to 220,575) that would include both symptomatic and asymptomatic individuals. It may also enter the exponential growth phase and it would become very difficult to contain at that stage. The situation can still be controlled if R_t can be brought down close to one. This indicates the need for more effective strategies and ensuring optimum testing to avoid underestimation of danger. However, it has to be noted that our study does not predict the β for India in the future, rather we extrapolated based on the range at the time of study.

In a modeling study of the Indian scenario, it was shown that India will have significant number of cases without intervention with widespread screening such as in airports delaying the epidemic and quarantine would slow local spread, decreasing the cumulative incidence by 2% to 62%[29]. Our study also shows how without decreasing R_t, India will have very high number of cases. Modeling studies have also been done that show that as the lockdown is eased, the testing should be increased to control the epidemic in India effectively, and the easing should be a gradual process and if lockdown is effective the peak prevalence can be decreased to a large extent[30].

Future research is needed to support the current findings that unlike other countries, in India, COVID-19 is mostly affecting

the 20–39 years age group since our data analysis was restricted to individuals in whom data was available. This study was conducted using data from the initial part of the epidemic in India. The SIR model used does not account for age-structure or comorbidities. Hence we are addressing more dynamicity and realness by taking into account multiple compartments with robust mathematical assumptions in our future models to study the current pandemic. Comprehensive studies based on the clinical manifestations and laboratory parameters including genomic sequencing to detect mutations of the virus also needs to be done for the Indian population to see if the clinical features and the viral strains differ from other populations.

To conclude, the COVID-19 epidemic in India is more affecting younger age groups as compared to other countries. Our mathematical model predicts that India will have a significant number of infected persons after the 40 days lockdown. Hence, strict social distancing, and optimum testing followed by isolation and quarantine are vital elements to control the COVID-19 epidemic.

Data availability
Underlying data
Raw data for daily cases of COVID-19 with gender, age and patient status obtained from: patientdb.covid19india.org

Raw data for estimation of β obtained from: https://www.statista.com/statistics/1104054/india-coronavirus-covid-19-daily-confirmed-recovered-death-cases/

Figshare:Mazumdaer A *et al.* study - COVID 19 India data, https://doi.org/10.6084/m9.figshare.c.4940472.v2[31].

This project contains the following data:
- Confirmed daily cases of COVID-19

- Full patient dataset COVID-19

- Indian population data

Data are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

Disclaimer: The views represented here are those of the authors and do not represent the institutions they are affiliated to.

References

1. Zhu N, Zhang D, Wang W, *et al.*: **A Novel Coronavirus from Patients with Pneumonia in China, 2019.** *N Engl J Med.* 2020; **382**(8): 727–733.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

2. Tang X, Wu C, Li X, *et al.*: **On the origin and continuing evolution of SARS-CoV-2.** *Natl Sci Rev.* 2020; nwaa036.
    Publisher Full Text | **Free Full Text**

3. World Health Organization: **WHO Director-General's opening remarks at the media briefing on COVID-19 - 18 March 2020.** 2020.
    Reference Source

4. Guan WJ, Ni ZY, Hu Y, *et al.*: **Clinical Characteristics of Coronavirus Disease 2019 in China.** *N Engl J Med.* 2020.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

5. **Modes of transmission of virus causing COVID-19: implications for IPC precaution recommendations.** 2020.
    Reference Source

6. van Doremalen N, Bushmaker T, Morris DH, *et al.*: **Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1.** *N Engl J Med.* 2020; **382**(16): 1564–1567.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

7. Lu S, Lin J, Zhang Z, *et al.*: **Alert for non-respiratory symptoms of Coronavirus Disease 2019 (COVID-19) patients in epidemic period: A case report of familial cluster with three asymptomatic COVID-19 patients.** *J Med Virol.* 2020.
    PubMed Abstract | **Publisher Full Text**

8. Rothe C, Schunk M, Sothmann P, *et al.*: **Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany.** *N Engl J Med.* 2020; **382**(10): 970–971.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

9. Qian G, Yang N, Ma AHY, *et al.*: **A COVID-19 Transmission within a family cluster by presymptomatic infectors in China.** *Clin Infect Dis.* 2020; pii: ciaa316.
    PubMed Abstract | **Publisher Full Text**

10. Chan JF, Yuan S, Kok KH, *et al.*: **A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster.** *Lancet.* 2020; **395**(10223): 514–523.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

11. Li Q, Guan X, Wu P, *et al.*: **Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia.** *N Engl J Med.* 2020; **382**(13): 1199–1207.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

12. Young BE, Ong SWX, Kalimuddin S, *et al.*: **Epidemiologic Features and Clinical Course of Patients Infected With SARS-CoV-2 in Singapore.** *JAMA.* 2020.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

13. Wu Z, McGoogan JM: **Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention.** *JAMA.* 2020.
    PubMed Abstract | **Publisher Full Text**

14. **Adaptive COVID-19 Treatment Trial (ACTT) - Full Text View - ClinicalTrials.gov.** 2020.
    Reference Source

15. **COVID 100casesIndia.** 2020.
    Reference Source

16. **Global surveillance for COVID-19 caused by human infection with COVID-19 virus.** *Interim Guidance.* WHO; 2020.
    Reference Source

17. Gupta N, Praharaj I, Bhatnagar T, *et al.*: **Severe acute respiratory illness surveillance for coronavirus disease 2019, India, 2020.** *Indian J Med Res.* 2020.
    PubMed Abstract | **Publisher Full Text**

18. **Predictive models for COVID-19-related deaths and infections.** *The Union.* 2020.
    Reference Source

19. Kermack WO, McKendrick AG, Walker GT: **A contribution to the mathematical theory of epidemics.** *Proc R SocLond Ser Contain Pap Math Phys Character.* 1927; **115**(722): 700–721.
    Publisher Full Text

20. Huppert A, Katriel G: **Mathematical modelling and prediction in infectious disease epidemiology.** *Clin Microbiol Infect.* 2013; **19**(11): 999–1005.
    PubMed Abstract | **Publisher Full Text**

21. **India - daily cumulative COVID-19 by type 2020.** 2020.
    Reference Source

22. Huang C, Wang Y, Li X, *et al.*: **Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China.** *Lancet.* 2020; **395**(10223): 497–506.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

23. Chen N, Zhou M, Dong X, *et al.*: **Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study.** *Lancet.* 2020; **395**(10223): 507–513.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

24. **India Tourism Source (Domestic & Outbound) Market Insights 2019 - ResearchAndMarkets.com.** 2019.
    Reference Source

25. Livingston E, Bucher K: **Coronavirus Disease 2019 (COVID-19) in Italy.** *JAMA.* 2020; **323**(14): 1335.
    Publisher Full Text

26. Huang Y, Yang L, Dai H, *et al.*: **Epidemic situation and forecasting of COVID-19 in and outside China.** *Bull World Health Organ.* 2020.
    Publisher Full Text

27. Philip SA: **How Tablighi Jamaat emerged as the 'largest known' Covid-19 source in South Asia.** 2020.
    Reference Source

28. **Huge crowds as lockdown sparks mass migration.** 2020.
    Reference Source

29. Mandal S, Bhatnagar T, Arinaminpathy N, *et al.*: **Prudent public health intervention strategies to control the coronavirus disease 2019 transmission in India: A mathematical model-based approach.** *Indian J Med Res.* 2020; **151**(2 & 3): 190–9.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

30. Gupta M, Mohanta SS, Rao A, *et al.*: **Transmission dynamics of the COVID-19 epidemic in India and modelling optimal lockdown exit strategies.** *medRxiv.* 2020.
    Publisher Full Text

31. Shewade HD: **Mazumdaer A et al study - COVID 19 India data.** *figshare.* Collection. 2020.
    http://www.doi.org/10.6084/m9.figshare.c.4940472.v2

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 1

Reviewer Report 29 May 2020

✔ **Soumya Swaroop Sahoo**

Department of Community & Family Medicine, All India Institute of Medical Sciences, Bathinda, Punjab, India

A very good attempt to present the Indian COVID-19 scenario by the authors. I congratulate them on their work.
However a few queries:

1. The data analysis has been performed on 1161 patients. To project it for such a large population has limited scope.

2. The range of the projection of 1.10 lakh to 2.20 lakh is quite wide. How do we interpret it?

3. In COVID, most of the patients recover in due course. If possible, the SEIR model could have been used for a better picture.

4. The discussion section contains a comparison with the descriptive data of different countries. It could have been compared with the modeling data of different studies of other countries to highlight the similarity/difference of the Indian scenario.

5. The authors in the discussion state that "we observed that the highest number of patients is in the 60–79 years age group category. The exact percentage projected needs to be mentioned.

6. The assumptions that the authors took for the modeling approach needs to be clearly stated in the methodology apart from birth and death rate.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Community Medicine and Epidemiology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 12 Jun 2020

PRIYAMADHABA BEHERA, All India Institute of Medical Sciences, Raebareli, India

Reply to Dr. Palepu's (Reviewer 1) comments:

Comment 1: The mathematical model used is "SIR", standing for "Susceptible, Infectious, Recovered/dead". This needs modification throughout the manuscript.

Author Response 1: We have used the full form as Susceptible-Infectious-Removed model, where removed includes Deceased and Recovered. Thank you.

Comment 2: In COVID-19, re-infection is being seen in a few patients. This can be explained in limitation as to why the "SIRS" model was not considered over "SIR". And also, not all susceptible individuals are infected. Why was the "SEIR" model not considered by authors?

Author Response 2: We have added the reason for not using the SIRS model in the methods section. We didn't consider the SEIR model because, at the time when this study was done, sufficient data was not available in the public domain for a reliable estimate of alpha for India which is the rate of transition from the exposed pool to the infected pool. In view of the limitations in this model, we have developed another model that is available at https://doi.org/10.1101/2020.05.13.20096826. Thank you.

Comment 3: Although, the model mentions that the birth rate and death rate were not taken into account, its effect on outcome estimation has not been explicitly mentioned. The reason that the duration of infection is much lesser than the lifetime of an individual can be explanatory for non-inclusion of birth and death rate.

Author Response 3: We have added the reason for the non-inclusion of vital dynamics in the methods section and how their inclusion would not significantly affect the results. Thank you.

Comment 4: Clear mention of the force of infection with proper notation is required.

Author Response 4: For the study, we have calculated gamma and tried to compare the real trends with trends for the varying value of beta. For this reason, we have not clearly mentioned the force of infection as it would not be reliably derived from our study. Also, because of varying R_t, it would be difficult to give a definite value for the force of infection in India. Thank you.

Comment 5: Reasons for the younger age group being affected more can be explained in greater detail.

Author Response 5: We have explained possible reasons like a high percentage of the population in the younger age group in India and possible younger age of imported cases in more detail in the discussion. Thank you.

Comment 6: Various studies have been published to date with different modeling techniques for the estimation of COVID-19 cases. They can be used for comparison in the discussion section.

Author Response 6: We have added discussions about other modeling studies of India in the discussion section. Thank you. When we were doing the study, the epidemic had progressed in other countries, so we compared our data with the actual data of the other countries. Also, the aim of our study was not to exactly predict the number of cases but to see the trend of the real data and get an idea about parameters like R_t, and contact rate. We extrapolated to give an idea of the number of cases India could have reached if the R_t had not been decreased by strict interventions. It must be noted that our study does not predict the future trend of beta, rather extrapolation has been done based on the range of beta in India till the time of the study.

Competing Interests: No competing interests were disclosed.

Author Response 12 Jun 2020

PRIYAMADHABA BEHERA, All India Institute of Medical Sciences, Raebareli, India

Reply to Dr. Sahoo's (Reviewer 2) comments:

Comment 1: The data analysis has been performed on 1161 patients. To project it for such a large population has limited scope.

Author Response 1: Thanks for your valuable comments. The data for 1161 patients were used for the study of the epidemiological features of the affected people in India. For the model, the real data of the number of active cases was plotted. However, we agree that using data from the initial stage of the epidemic in India is one of our limitations, and this has been included in the discussion. This study was done to provide an overview of what India might be dealing with.

Comment 2: The range of the projection of 1.10 lakh to 2.20 lakh is quite wide. How do we interpret it?

Author Response 2: We have incorporated the explanation for this wide range of finding in the discussion. With the preliminary data available in the public domain, we obtained this interval based on the range of beta at the time of the study. Our study does not predict the beta for India in the future, rather we extrapolated based on the range at the time of the study. The range gives an

idea of the number of infections if the contact rate wasn't affected by lockdown and strict social distancing. Thank you.

Comment 3: In COVID, most of the patients recover in due course. If possible, the SEIR model could have been used for a better picture.

Author Response: As we have mentioned in the reply to Dr. Palepu's 2nd comment, we didn't consider SEIR model because, at the time when this study was done, sufficient data was not available in the public domain for a reliable estimate of alpha for India which is the rate of transition from the exposed pool to the infected pool. In view of the limitations in this model, we have developed another model that is available at https://doi.org/10.1101/2020.05.13.20096826. Thank you.

Comment 4: The discussion section contains a comparison with the descriptive data of different countries. It could have been compared with the modeling data of different studies of other countries to highlight the similarity/difference of the Indian scenario.

Author Response 4: When we were doing the study, the epidemic had progressed in other countries, so we found it better to compare our data with the actual data of the other countries. We have modified our discussion including modeling studies form the Indian scenario. Thank you.

Comment 5: The authors in the discussion state that "we observed that the highest number of patients is in the 60–79 years age group category. The exact percentage projected needs to be mentioned.

Author Response 5: This observation is based on the normalized values calculated and tabulated in Table 1. The language has been modified for clarity. Thank you.

Comment 6: The assumptions that the authors took for the modeling approach needs to be clearly stated in the methodology apart from birth and death rate.

Author Response 6: Apart from the assumption regarding birth and death rate, we have assumed the value of removal rate gamma to be constant as we have mentioned in the text. We assume uniform transmission for all our extrapolations. These have been described in the method section. Thank you.

Competing Interests: No competing interests were disclosed.

Reviewer Report 22 May 2020

https://doi.org/10.5256/f1000research.25930.r63079

✔

Sarika Palepu iD

Department of Community and Family Medicine, Vir Chandra Singh Garhwali Government Institute of Medical Sciences and Research, Srinagar, Uttarakhand, India

1. The mathematical model used is "SIR", standing for "Susceptible, Infectious, Recovered/dead". This needs modification throughout the manuscript.

2. In COVID-19, re-infection is being seen in a few patients. This can be explained in limitation as to why "SIRS" model was not considered over "SIR". And also, not all susceptible individuals are infected. Why was the "SEIR" model not considered by authors?

3. Although, the model mentions that the birth rate and death rate were not taken into account, its effect on outcome estimation has not been explicitly mentioned. The reason that the duration of infection is much lesser than the lifetime of an individual can be explanatory for non-inclusion of birth and death rate.

4. Clear mention of the force of infection with proper notation is required.

5. Reasons for the younger age group being affected more can be explained in greater detail.

6. Various studies have been published to date with different modelling techniques for the estimation of COVID-19 cases. They can be used for comparison in the discussion section.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Public health and Epidemiology of infectious diseases

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias

- You can publish traditional articles, null/negative results, case reports, data notes and more

- The peer review process is transparent and collaborative

- Your article is indexed in PubMed after passing peer review

- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research